Neural signatures underlying individual differences in source monitoring abilities

Øivind Solberg
Masters Degree in Psychology
Department of Psychology
University of Oslo, 2007

Author: Øivind Solberg

Title: Neural signatures underlying individual differences in source monitoring abilities

Supervisor: Tim Brennen

Abstract

In the present study, source-monitoring processes that are required to distinguish a memory trace of a true event from an internally generated false memory were investigated, using the Deese-Roediger-McDermott paradigm (DRM) to induce false memories, while obtaining fMRI measurements. In order to explore individual differences in memory performance and source monitoring abilities, participants were divided into Low False Memory (LFM) and High False Memory (HFM) groups based on accuracy for critical lure words. Subsequent analyses of the two groups' behavioural data revealed intrinsic differences in accuracy and response time patterns, while post hoc ROI analyses of the groups' functional data, consistent with the behavioural findings, revealed significant enhanced activations during recognition for the LFM group compared to the HFM group in areas previously shown to be linked to memory and source monitoring performance. As far as we know, this is the first study to identify possible neural mechanisms underlying individual differences in source monitoring abilities related to false memory susceptibility.

Key words: False memory, fMRI, medial temporal lobe, dorsolateral prefrontal cortex, source monitoring, individual differences, DRM, recognition

Preface

Working in a new and explorative field within psychological research has given me many new insights and great challenges during the year I have been working with this thesis. The preparations prior to my four month stay in Maastricht, Netherlands were demanding, and like any other large changes in life, a bit chaotic. Still, my endeavor into the realm of cognitive neuroscience and fMRI research has given me great insights into a field within psychology which was fairly unknown to me prior to the birth of the project.

Within the scope of the project I took part in all the stages of the study, i.e. fully participating in the development, planning and preparation of the study, including setting up the paradigm, preparing the stimuli, programming the stimulus presentation with E-prime, piloting prototype setups, recruitment, fMRI data collection, the preprocessing of the MR data and the statistical analysis of the final results. My involvement in every aspect of the project, the execution of the study abroad, and the final write-up of the thesis has been an enriching learning experience which will be of great use to me in the years to come and I will never forget it.

Still, I am indebted to a number of people, departments and organisations for various kinds of support that made the project possible. Some I will mention here, but I already wish to apologise to those of you who are not mentioned by name or forgotten in the hectic chaos that usually erupts prior to a master thesis deadline.

First and foremost I wish to express my gratitude to my supervisor, colleague, mentor and friend Professor Tim Brennen for his never-ending belief in the project and my abilities as a master student. Tim guided me through an extraordinary demanding year with extreme time constraints and relentless hard work in a flawless manner always expressing trust, but also demanding results. Special thanks for going beyond all requirements set for supervision by always taking the time to give me feedback, comments and that last “push” in order to make me reach the required milestones. Also big thanks for all the great laughs we had during the year, without which I wouldn’t have managed. I’m forever grateful.

Secondly, I wish to thank Drs. Bettina Sorger for her excellent supervision during my stay in Maastricht. Bettina taught me everything I know about fMRI and the BrainVoyager software, while we worked side by side through endless preprocessing and analysis sessions in the computer lab and fMRI scanner room. Special thanks for coming to Oslo during the Easter holiday in order to complete the analysis. I could not have done it without you.

Thirdly, I would like to express special thanks to master student Linsey Raymaekers for all the support and help she offered me during my stay in Maastricht. Linsey assisted me during all the stages of the study, but most of all she offered me deeply needed empathic support and endless energy input during the hard last month in Maastricht. Thanks, I will not forget it.

Fourthly, I wish to thank Dr. Jochen Seitz for his participation in the false memory team and for all the help he gave me during the different stages of the project. Special thanks for proofing the imaging results together with Bettina.

Huge thanks to Phd. student Kjell Ivar Øvergård for valuable lessons in SPSS and help during the statistical analysis of the behavioural data, and to Dr. Elke Geraerts and Professor Rainer Goebel for setting up and managing the project in Maastricht together with Professor Tim Brennen.

I would also like to thank the Faculty of Psychology in Maastricht and the Department of Psychology in Oslo, which provided me with excellent working conditions during my stay in the Netherlands. The project received financial support from the Department of Psychology in Oslo, the ERASMUS program for exchange students, and was awarded a departmental "studiestipend", without which the study could not have been carried out.

Øivind Solberg, Oslo, 2007

Introduction

Memory is rarely perfect and its imperfections are not only limited to simple forgetting. We know from empirical studies that our memories can be lost, drastically changed or even be false, and still remain as compelling and 'real' as accurate memories (e.g. Roediger & McDermott 1995; Hyman & Pentland, 1996; Lindsay, Hagen, Read, Wade & Garry, 2004). Furthermore, we often do not remember the actual details of our experiences accurately, and to fill in the missing information, we often assume, infer, or imagine the details of the content of our memories based on our own experience, emotions or other people's suggestions, constructing integrated memories of the events. However, these internally generated thoughts, ideas, associations or beliefs sometimes go beyond filling in minor gaps, and create false memories or memories of events that never happened (e.g. Roediger & McDermott 1995; Mather, Henkel & Johnson, 1997; Loftus, 2003). When misattributions or source monitoring errors like this occur, the retrieved information, such as an imagined episode, is assigned to the wrong source and thought to be real (Schacter & Slotnick, 2004).

"Source monitoring" in this article is defined as the individual ability to identify the source of remembered information, and will in particular distinguish between an external source (e.g. words presented on a screen), and an internal source (e.g. the participants themselves generating words). In the lab, this source monitoring ability can be investigated using a procedure introduced by Deese (1959) and subsequently developed by Roediger and McDermott (1995), dubbed the DRM paradigm. In this paradigm participants typically try to remember lists of words, with the members of each list being associatively related to a non-presented target word, the so-called ''critical lure'', prior to a recognition test. For example, a standardized list with the non-presented critical lure ''sleep'', begins ''bed, rest, awake, tired…etc'' (Stadler, Roediger, & McDermott, 1999). Since concepts related in semantic memory are linked in such a way that accessing one concept (e.g. *bed*) sends activation through linked pathways to related concepts (e.g. *sleep*), participants subsequently falsely recognize the nonpresented semantic associate lure of the lists, often at rates comparable to that of words that were actually presented. Moreover, when asked for a rating of how sure they were that the word was actually presented, it is evident that on a variety of different measures (Remember/Know, confidence etc.), that the ''memory'' for the non-event is compelling (e.g. Tulving, 1985; Gardiner & Java, 1993, Schacter, Norman & Koutstaal, 1998). This false memory phenomenon is thought to be induced by semantic elaboration and implicit associative responses to the highly associative stimuli, to which people overtly or

covertly generate the nonpresented critical lure (McDermott & Watson, 2001). An essential feature of this semantic elaboration process is the integration of incoming information with preexisting semantic knowledge (Kim & Cabeza, 2006). The integration process strengthens the formulation of veridical memory traces, but also contributes to the creation of false memories or illusory memory traces. False memories induced by the DRM paradigm are therefore seen as source monitoring errors (see e.g., Brédart, 2000; Schacter, Verfaellie, & Pradere, 1996) and are though to occur due to a breakdown in source monitoring systems that differentiate the activation of internally generated concepts from representations of previously studied words.

In 2001, Cabeza and colleagues examined this idea by investigating the neural patterns associated with true and false recognition using a modified DRM paradigm and fMRI measurements (Cabeza, Rao, Wagner, Mayer & Schacter 2001). They hypothesized that, during recognition, previously studied words, but not semantically associated critical lures, would activate regions initially involved in encoding perceptual source information, whereas regions involved in the encoding and retrieval of semantic information would show comparable activation during both true and false recognition. Twelve participants watched a videotape segment in which a male and a female speaker alternatively presented lists of associated words. Participants were instructed to remember not only the presented words, but also which of the speakers who presented them. The participants then performed a recognition test including words presented in the study lists (true items), new words closely related to studied words (critical lures), and new unrelated words (new items).

Results revealed a dissociation between two medial temporal lobe (MTL) regions in which hippocampus was similarly activated for true words and critical lures, suggesting the recovery of semantic information, whereas the parahippocampal gyrus was more activated for true words compared to new non studied words and critical lures, suggesting greater recovery of perceptual or contextual information during true compared to false recognition. The study also yielded a dissociation between two prefrontal cortex regions in which the bilateral dorsolateral prefrontal cortex was more activated for true words and critical lures compared to new words, possibly reflecting source monitoring of retrieved information, whereas left ventrolateral prefrontal cortex was more activated for new words compared to true words and critical lures, possibly reflecting semantic processing.

A later study by Okado and Stark (2003) revealed similar results. In this study the authors used fMRI to compare true recognition of previously perceived events with false recognition of previously imagined events. The experiment was divided into three phases; a

study phase, a misinformation task, and a final recognition test. In the study phase, prior to scanning, words naming concrete objects were presented auditorily and the participants were asked to imagine the object. An actual picture of the object followed the presentation of half of the objects. A misinformation task was employed between the study phase and the test phase, in an effort to increase the number of false memories. During the misinformation task, the participants were given a version of the test phase in which they were encouraged to lie about seeing a picture that they had only imagined during the previous study phase Later, during the "real" test phase, the lying supposedly induced a misattribution error, which served as a self-generated source of misinformation. In the recognition phase, which was the only in-scanner portion of the experiment, the participants heard names of objects they had seen accompanied by a picture, objects that were presented without a picture, and new objects. The participants were asked to determine if they had actually seen a picture of the object during the study phase or not.

Results revealed that occipital and right posterior parahippocampal gyrus showed greater activity for seen pictures than for imagined self generated pictures. In contrast, the right anterior cingulate regions showed the opposite pattern, suggesting a distinction between retrieval processes that yield true and false memories. Finally, the left parietal lobe, left frontal lobe, and areas within the MTL showed similar activity for seen pictures and for self generated imagined pictures believed to be seen, suggesting an inability to detect differences between retrieval processes that yield true memories and retrieval processes that yield false memories. Thus, the results of the study suggested that for true and false memories of pictures, the occipital and posterior parahippocampal regions show activity that distinguishes these memories, whereas the left frontal and parietal activity reflects how much both true and false memories are believed to be true (Okado & Stark, 2003).

More recently, Daselaar, Fleck and Cabeza (2006) examined the mechanisms underlying recognition and source monitoring abilities further within the context of a lexical decision task. In this study participants studied a list of normal words and unpronounceable non-words, one word at a time, prior to scanning. After a 30 min break the participants were scanned while performing a recognition task. During the recognition task participants viewed an equal amount of old words and new words while indicating old/new responses followed by a subjective confidence rating ranging from 1(low) to 4 (high).

Using a parametric approach based on an eight point oldness scale (confidence ratings combined with the OLD and NEW responses (1= definitely NEW to 8= definitely OLD)), the authors managed to isolate retrieval-related activity associated with recollection, familiarity,

and novelty within the MTL. Further, the authors identified a triple dissociation among the posterior half of the hippocampus, which was associated with recollection, the posterior parahippocampal gyrus, which was associated with familiarity, and the anterior half of the hippocampus and rhinal regions, which were associated with novelty. Multiple regression analyses based on individual trial activity also indicated that all three memory signals (recollection, familiarity, and novelty) made independent contributions to memory performance. Finally, contrary to Cabeza et al.'s findings (2001), a ventrolateral prefrontal cortex region was associated with recollection, whereas right dorsolateral prefrontal cortex showed novelty-related activity. In sum, the findings revealed that different brain regions can be differentially involved in recollection, familiarity, and novelty processes, possibly supporting a recollection/familiarity distinction.

Taken together, the findings of the aforementioned studies indicate that activity within hippocampus, parahippocampal regions and occipital regions distinguish true from false memories, whereas left frontal and parietal activity reflects how much both true and false memories are believed to be true. Furthermore, the findings indicate that activity within dorsolateral prefrontal cortex reflects source monitoring of retrieved information and that activity within ventrolateral prefrontal cortex reflects semantic processing of novel information. However, a more critical reading of the different results reveals a less clear cut picture.

Although reporting areas of interest in relation to source monitoring, Cabeza et al. (2001) regrettably did not measure individual variations in the data and the corresponding neural activity that could possibly reflect differences in source-monitoring abilities across participants. Further, the analysis of variance (ANOVA) reported were conducted separately for each of the functional images, followed by posthoc pairwise comparison of the three conditions (true, critical lure and new). The significant results reported in hippocampus and parahippocampal gyrus were henceforth produced by conducting separate ANOVA's on the average hemodynamic response functions (HRF) from images showing maximal differences, primarily image 2 and 3, whereas results derived from image 1, 4 and 5 were not reported. This type of statistical analysis makes the results difficult to interpret since contemporary functional imaging software packages, such as BrainVoyager QX and SPM2, run the ANOVA's and the pairwise comparisons for all the obtained images, making the statistical tests more conservative in relation to levels of significance. The illustrations presented in the Cabeza et al. article are also difficult to interpret. The illustrated HRF's were smoothed using a special feature in Microsoft EXCEL and were therefore presented without error bars,

showing an unnatural "slope down", i.e. a rapid decrease in the HRF signal towards the end of the trial. Taken together, these limitations make the findings of the study difficult to interpret.

In the Okado and Stark (2003) study, the authors note that although they found activity in the right parahippocampal gyrus which extended to include some of the right parahippocampal cortex, the activity was largely posterior to what one normally defines as parahippocampal regions, and that the pattern of activity was strikingly similar to the pattern observed in occipital regions. Although the MTL usually is associated with memory encoding and retrieval processes (see, e.g., Henson, 2005 for review), the study did not observe significant activity throughout the MTL regions that differentiated the different trial types. Okado and Stark explain their null finding of hippocampal and other MTL activity by referring to confounding effects such as incidental encoding and/or the presence of episodic or source memory components in all the trial types analyzed. According to the authors, such commonality may have resulted in similar levels of activity across the trial conditions in many of the MTL regions, ultimately flattening the activation pattern when the contrasts were applied. Regrettably, the study failed to report individual data which could have shed more light on these findings, possibly showing that high intra-variability in the participant's MTL activity could better explain the puzzling results.

In the Daselaar et al. (2006) study, the authors note difficulties in identifying the different regions of interest within the MTL without the use of an unconventionally low threshold. In fact, the authors report that the anterior MTL regions they observed would not have survived the threshold conventionally used in event-related fMRI studies. Here it is worth noting that different authors use different statistical criteria for "reliable activations" and that precise localization within MTL is rarely achieved, given the susceptibility-induced distortions associated with echo planar fMRI, and the fact that most authors report mean locations within normalized brains and/ or template brains, which reduce spatial resolution (Henson, 2005). Current methods of matching different brains therefore do not eliminate individual differences in functional areas across participants since the positions of these areas are not well predicted by the gross anatomical landmarks used in many regions of the brain (Brett, Johnsrude and Owen, 2002). These problems therefore cause uncertainty in localization, especially for brain areas involved in higher cognitive function and could explain many of the deviating findings within the fMRI literature.

This being said, Henson (2005) note that even though the fMRI literature on memory retrieval is numerously marked by failures to observe MTL activity, a trend seems to emerge indicating that bilateral hippocampus and posterior regions of the medial temporal cortex, e.g.

parahippocampal regions, appear to be particular important for retrieving source information and associations between distinct items in memory. Although further research is needed, findings within the false memory literature and the aforementioned studies also generally confirms this trend, suggesting that hippocampal, parahippocampal, occipital and dorsolateral prefrontal regions play an important role when distinguishing true from false memories.

The present study was designed to further explore the false memory phenomena observed in the DRM paradigm using fMRI measurements during encoding and recognition. A novelty with the study was the concept of dividing participants into high and low false memory groups based on their behavioural performance in order to disentangle possible individual differences in neural activity related to source monitoring and memory performance. We hypothesized that participants' response patterns during recognition in the DRM paradigm, especially for critical lure words, would reflect individual source monitoring abilities, and that a comparative analysis would uncover distinct activity variations, especially in dorsolateral prefrontal cortex, hippocampus and parahippocampal cortices, possibly disentangling some of the deviating findings described earlier in the introduction.

Method

Participants

Thirty-five participants were recruited from the University of Maastricht community and administered an fMRI pre-scan security check questionnaire. Fifteen of the thirty-five participants were excluded for criteria such as near-or farsightedness, claustrophobia, tattoos and/or irremovable metal objects in the body. Twenty healthy right handed participants were then authorized for scanning and invited to a pre-scan screening phase.

In the screening phase participants were asked to fill in five different self report scales; the Beck Depression Inventory (Beck and Steer, 1987), the Childhood Trauma Questionnaire (Bernstein, Stein, Newcomb, Walker, Pogge, Ahluvalia, Stokes, Handelsman, Medrano, Desmond, & Zule, 2003), the Dissociative Experiences Scale (Bernstein & Putman, 1986), the Creative Experiences Questionnaire (Merckelbach, Horselenberg & Muris, 2001) and the Cognitive Failures Questionnaire (Broadbent, Cooper, Fitzgerald & Parkes, 1982). Since none of the invited participants had high scores or scores above the clinical threshold on any of the self report scales, ten participants were randomly selected from the group. All of the ten participants (6 female & 4 male, average age of 22 years, $SD = 2.8$) gave written informed consent prior to participation and the fMRI sessions were conducted under a protocol approved by the Maastricht University Faculty of Psychology Ethics Committee.

Materials

The material used in the present DRM paradigm consisted of twenty-six Dutch lists consisting of fifteen semantic associate words related to twenty-six critical lures (a total of 390 words, see appendix for list of critical lures). The lists were based on material previously employed in the Maastricht lab by Geraerts, Smeets, Jelicic, Heerden and Merckelbach (2005). In preparation of the lists, pilot work showed that the lists produced rates of false recognition comparable to those reported by Stadler, Roediger and McDermott (1999).

Procedure

The overall procedure used in the present study was similar to that of the typical DRM paradigm (e.g. Roediger and McDermott, 1995), partly modified in order to better suit an fMRI study. The whole study was conducted inside the scanner, and, in order to reduce strain on the participants, the study was divided in two sessions, each consisting of an encoding run and a recognition run, giving four runs in total. Before participating in the actual scanning session, all participants took part in a "dummy scan" session where they performed brief versions of the encoding and recognition tasks. During this practice session participants were familiarized with the pressing of the different buttons and learned the response mapping by heart in order to minimize differences in reaction times between the different fingers and any impact of response mapping errors on the imaging results.

In each of the two encoding runs, participants were individually instructed to try to remember 13 of the 26 lists in preparation for a subsequent memory test. Participants viewed 195 words per run at a rate of 1500 ms per word with a cross-hair fixation pause of 1000 ms between words. The words were presented one by one, sequentially within each list. Each list-block had a duration of 35700 ms and was followed by a cross-hair pause for 20000 ms. Each encoding run was followed by a distracter phase of 15 min[1]. During the distracter phase anatomical scans were conducted while participants viewed a cartoon.

After each distracter phase, a recognition run corresponding to the recently completed encoding run began. Each run began by instructing participants to judge the relative extent to which they consciously recollected the experience of seeing the words presented in the encoding run, using the Remember/Know/New distinction believed to assess the phenomenological qualities of memories (Tulving, 1985; Gardiner & Java, 1993; Gonsalves,

[1] Since the main topic of this article is source monitoring during recognition, the fMRI data and results from the encoding runs will only be presented briefly in the discussion section.

Kahn, Curran, Norman, & Wagner, 2005). Participants were instructed to use "Remember" judgments for vivid, consciously remembered memories, and "Know" judgments for confident memories for which they could retrieve no associated specific detail. New responses were to be used for words believed to be novel. Each of the words was displayed for 2500 ms on the screen, immediately followed by the presentation of a cross-hair fixation for 10000 ms. Each recognition run consisted of 91 words corresponding to the 13 lists presented in the preceding encoding run (3 OLD words, 3 NEW words and 1 criticallure word per list x 13 lists = 91 words). For each trial participants were instructed to respond within 2500 ms by pressing buttons corresponding to their right index finger, right middle finger or right ring finger using a fiber optic response box. An index finger response indicated a ''OLD Remember'' response, (corresponding to a recollection of the word), a middle finger response indicated a ''OLD Know'' response, (corresponding to a memory without specific detail), and finally, a ring finger response indicated a ''NEW or novel'' response for words participants hadn't seen before.

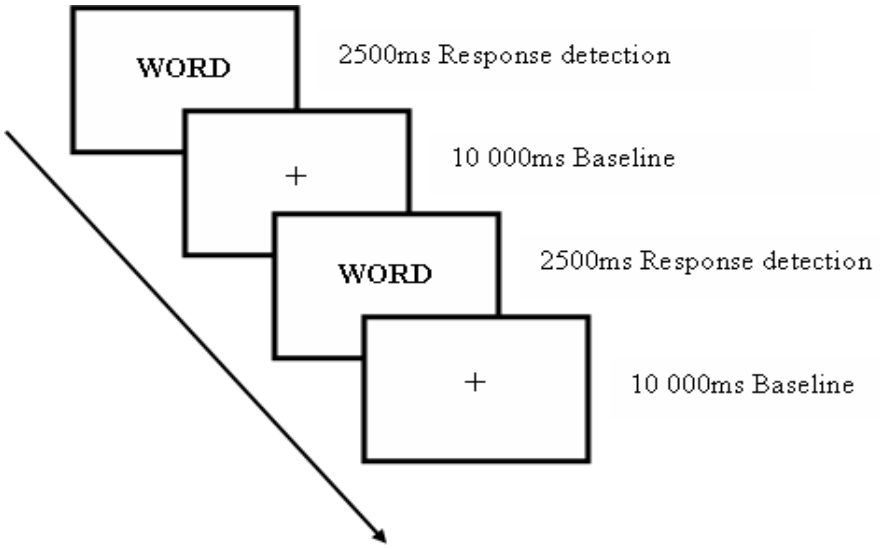

Figure 1. *Schematic diagram of recognition trials. Each word was presented for 2500 ms, followed by a 10 000 ms fixation period. Participants were asked to respond within the 2500 ms timeframe and then fixate on the crosshair while waiting for the next word presentation.*

Image acquisition and data analysis

A Siemens 3 T Magnetom Allegra head scanner unit was used to acquire both high-resolution anatomical and functional MR images using a standard volume coil. For each participant, a three-dimensional (3D) T1-weighted data set encompassing the whole brain was acquired after initial positioning scout images had been obtained. In order to ensure precise anatomical reference, a Modified Driven Equilibrium Fourier Transform (MDEFT) sequence

(scan parameters: repetition time [TR] =7.92 ms, echo time [TE]=2.4 ms, flip angle [FA]=15°, field of view [FOV]=256×256 mm2, matrix size=256×256, number of slices=176, slice thickness=1 mm, no gap, total scan time=13 min and 43 s) was used. This sequence facilitates later cortex segmentation and spatial localization, especially important when working within higher cognitive level regions. Functional measurements were acquired using a gradient-echo planar imaging EPI sequence in a partial field of interest positioned brain coverage (scan parameters: TR = 1250 ms, TE = 25 ms, FA = 73º FOV = 192 × 192 mm² and a matrix size of 64 × 64, number of slices= 23, slice thickness 3mm, distance factor 33%). Each encoding run lasted 14.43 min corresponding to 706 volumes per run, while each recognition run lasted 19.18 min corresponding to 926 volumes per run. The participants' recognition task begun in synchrony with the acquisition of the seventeenth volume, using the scanner pulse to synchronize stimuli presentation with the slice acquisition.

The analysis of the anatomical and functional data sets was performed using BrainVoyager QX (Version 1.8; Brain Innovation, Maastricht, The Netherlands). Prior to the regression analysis, the functional data sets were subjected to a series of preprocessing operations. To exclude scanner-related signal drifts, a linear trend removal was performed. Further, temporal high-pass filtering was applied to remove temporal frequencies lower than 3 cycles per run. Small interscan head movements, which altogether not exceeded a translation of 3 mm, were corrected for by a sinc-interpolated, rigid body algorithm rotating and translating each functional volume in 3D space. The functional data were then smoothed spatially with a Gaussian kernel of 4mm full-width half-maximum (FWHM). To enable the comparison between subjects, all anatomical as well as the functional volumes were spatially normalized in Talairach space (Talairach transformation; Talairach and Tournoux, 1988). In order to optimize visualization of the data, general functional networks were projected on an individual participant's right and left cortical flat maps. The cortical flatmaps were generated using a special cortex segmentation, inflation, and flattening procedure implemented in the BrainVoyager QX software (Goebel & Jansma, 2006).

The functional data were then analyzed using multiple regression models consisting of predictors which corresponded to the particular experimental conditions of the study. To compare the BOLD responses during the experimental conditions, fixed effects general linear models (GLM) contrasts were computed. If not stated differently, signal differences with a threshold of $p<0.05$ (one-tailed, Bonferroni-corrected for multiple comparisons) were considered significant. In the regions-of-interest analyses (ROI analyses), no Bonferroni correction had to be performed.

Behavioural Results

General results

The mean accuracy and mean response times for recognition of previously studied old words, critical lure words and other new words were computed for each participant. A one-way repeated measures ANOVA with the factor of Wordtype (Old vs. New vs. Critical Lure) showed a significant main effect [$F(2, 18) = 25.469$, $p < .001$, $\eta_p^2 = .739$]. Participants responded above chance for OLD (66%) and NEW (78%) words, but showed a strong false memory effect, only correctly rejecting 30 percent of the critical lure words.

Consistent with previous reports, the mean response time for hits (1302ms) and correct rejections of new words (1296ms) were faster than response times for incorrectly identified critical lure words (1386ms). A paired samples t-test revealed a significantly longer response time for incorrectly identified critical lure words compared to hits [$t(9)=2.28$, $p < .05$], but not for the other comparisons; hits vs. correct rejections [$t(9)<1$, ns] and correct rejections vs. incorrectly identified critical lures [$t(9)=1.27$, ns].

Group results

In order to further explore and identify differences in source monitoring abilities both behaviorally and neurologically, the ten participants were subsequently divided into Low False memory (LFM, 3 female, Mean age= 20.8, SD=1.33) and High False memory (HFM, 3 female, Mean age=23.2, SD=3.31) groups based on their accuracy for the critical lure words. The mean number of correct responses for the different wordtypes is shown in table 1 for each group, with standard deviation in brackets.

Group		Mean Number (SD)	Minimum	Maximum
High False Memory	Critical lures	4.70 (2.81)	3	8
	NEW	54.40 (2.89)	50	58
	OLD	55.77 (13.10)	36	69
Low False Memory	Critical lures	11.13 (2.47)	10	15
	NEW	67.13 (3.99)	57	68
	OLD	46.78 (12.58)	36	66

Table 1. *The mean number of correct responses for the different wordtypes by group, with standard deviation in brackets. Note that the number of times a person incorrectly "recognized" a critical lure word is 26 minus the number in the table. Each participant saw 26 critical lure words, 78 new words and 78 old words.*

A mixed two-way ANOVA with a between subjects factor of Group (LFM & HFM) and a within factor of Wordtype (OLD, NEW & CL) was carried out with the dependent variable of accuracy. It showed a significant main effect of Group [$F(1,8) = 8.731$, $p < .018$, $\eta_p^2 = .522$], a significant main effect of Wordtype [$F(2,16) = 38.997$, $p < .001$, $\eta_p^2 = .830$] and a significant interaction effect for Group and Wordtype [$F(2,16) = 5.782$, $p < .013$, $\eta_p^2 = .420$]. Separate paired samples t-tests for each group were used in order to disentangle the interaction. For the HFM group it revealed significantly lower accuracy for critical lure words compared to for OLD words [$t(4)=5.06$, $p<0.01$], whereas the same comparison for the LFM group did not [$t(4)=1.79$, ns].

Descriptive statistics further revealed that the HFM group's number of "remember" responses were higher than the number of "know" responses for the critical lure words, while the opposite was true for the LFM group, indicated by a higher number of "know" responses for the critical lure words compared to "remember" responses, as shown in table 2.

Group		Mean Number (SD)	Minimum	Maximum
High False Memory	Remember	11.2 (5.93)	5	19
	Know	9.0 (4.58)	2	14
	New	4.6 (2.88)	1	8
Low False Memory	Remember	6.6 (1.52)	5	9
	Know	8.0 (2.83)	5	11
	New	11.0 (2.65)	8	15

Table 2. *The mean number of "Remember", "Know" and "New" responses for critical lure words by the High False Memory (HFM) and the Low False Memory (LFM) group.*

A mixed two-way ANOVA with the between subjects factor of Group (LFM & HFM) and the within subjects factor of wordtype accuracy (hits, correct rejections & false critical lures) was carried out with the dependent variable of response time. It showed a significant main effect of Wordtype [$F(2,16) = 3.74$, $p < 0.05$, $\eta_p^2 = .319$], and a significant interaction effect for Group and Wordtype [$F(2,16) = 11.30$, $p < .001$, $\eta_p^2 = .586$], whereas the main effect of Group was not significant [$F(1,8) < 1$, ns]. Separate paired samples t-tests for each group were used in order to disentangle the interaction. For the LFM group it revealed significantly

longer response times for false critical lures compared to hits, $t(4)=4.93$, $p<0.01$, whereas the same comparison for the HFM group did not [$t(4)<1$, ns] as shown in figure 2.

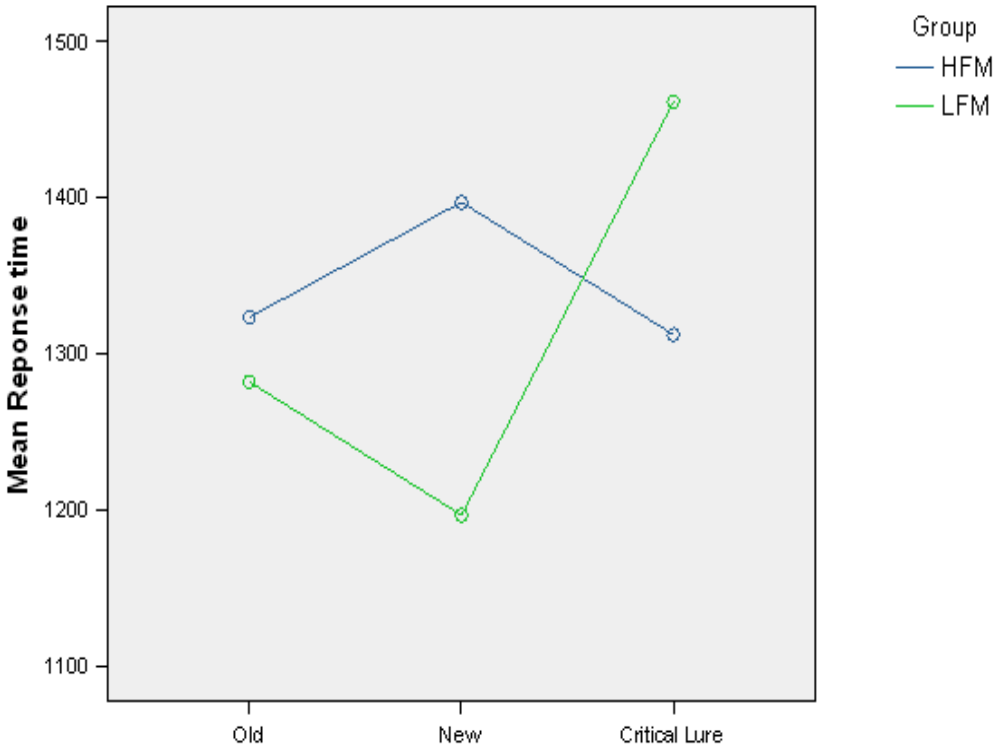

Figure 2. *Mean Response times (ms) for Old hits, correct rejections of New words and false identifications of critical lure words by each group.*

Imaging Results

General recognition network

First, general regions of brain activation associated with the overall effect of the recognition task were identified by pooling all the three trial types (previously studied words (OLD), novel words (NEW) and critical lure words (CL)) and comparing them to baseline fixation. Significantly activated regions (p<0.01, Bonferroni corrected) were revealed within parietal lobes, frontal lobes, occipital and temporal lobes as shown on left and right hemisphere flatmaps (fig. 3a and b).

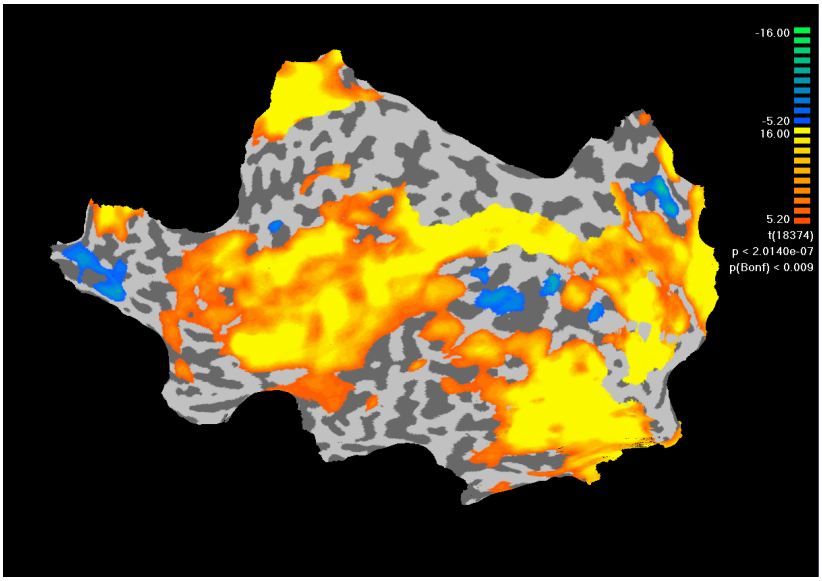

Figure 3a
General recognition networks for the left hemisphere projected on to an individual participant's flatmap. Only voxels with a significant activity of p<0.01Bonferroni corrected for whole brain multiple comparisons are shown.

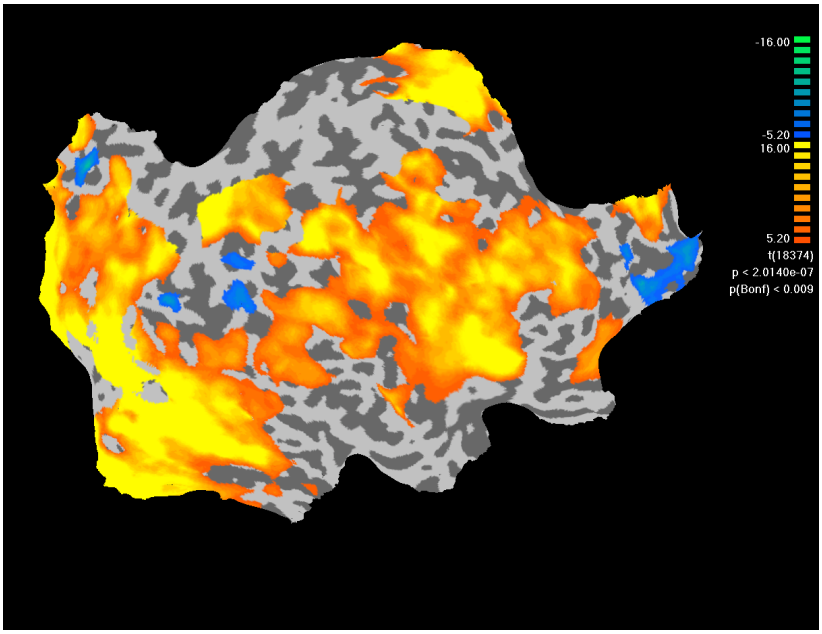

Figure 3b
General recognition networks for the right hemisphere projected on to an individual participant's flatmap. Only voxels with a significant activity of p<0.01Bonferroni corrected for whole brain multiple comparisons are shown.

Wordtype effects

Contrasts of "OLD vs. NEW" and "CL vs. OLD" was applied in order to identify regions that showed significant differential activation for the different wordtypes. Based on the contrasts, 10 regions of interest (ROIs) were identified in which we conducted separate ROI analyses. Each ROI and the corresponding result of the pairwise contrasts (OLD vs. NEW vs. CL) are shown in table 3.

Regions of interest	**(x,y,z)**	**Pairwise contrasts (beta weights)**
OLD vs. NEW contrast		
Left dorsolateral prefrontal cortex 1	-35,52,15	OLD(.457)~CL(.421) > NEW(.194)
Left dorsolateral prefrontal cortex 2	-43,41,18	OLD(.501)~CL(.429) > NEW(..227)
Right dorsolateral prefrontal cortex	37,53,1	CL(.260)~OLD(.239) > NEW(.034)
Left anterior hippocampus	-18,-21,-7	OLD(.508)~CL(.404) > NEW(.256)
Left posterior parahippocampal gyrus*	-12,-34,-11	OLD(.424) > NEW(.247)~CL(.194)
Left Medial Superior Frontal Gyrus 1	-2,38,29	CL(.534)~OLD(.482) >NEW(.336)
Left Medial Superior Frontal Gyrus 2	-2,19,46	CL(1.269) > OLD(1.156) > NEW(.980)
CL vs. OLD contrast		
Right middle frontal gyrus 1	49,18,35	CL(.976) > OLD(.675) > NEW(.539)
Right middle frontal gyrus 2	30,0,53	CL(.808) > OLD(.535) > NEW(.417)
Left middle frontal gyrus	-44,16,26	CL(.747) > OLD(.510)~ NEW(.431)

Table 3. *Neural regions identified applying "OLD vs. NEW" and "CL vs. OLD" contrasts. Results of the pairwise contrast within each of the ROIs are shown with beta weights in brackets. Talairach coordinates (x,y,z) refers to the center of activation within each region .*The left posterior parahippocampal ROI was identified using a non-significant threshold.*

Medial temporal lobe. In line with previous research, a dissociation within the medial temporal lobe (MTL), specifically between right anterior hippocampus and left posterior Parahippocampal gyrus, was identified. The hippocampal region was more activated for OLD and CL words than for NEW words, with no difference between OLD and CL words (Fig. 4a). Pairwise contrasts of the average BOLD signal within left hippocampus yielded significant differences between CL and NEW [$t(22)= 1.98$, $p< 0.05$] and between OLD and New [$t(22)= 4.74$, $p<$, 0.001], but not between OLD and CL [$t(22)=1.37$, ns]. A different pattern emerged in the left posterior parahippocampal gyrus which was more activated for

OLD than for CL and NEW words, with no difference between CL and New words (Fig. 4b). Pairwise contrasts of the average BOLD signal yielded significant differences between OLD and CL [$t(22)=2.26$, $p< 0.05$] and between OLD and NEW [$t(22)=2.46$, $p< 0.01$], but not between CL and NEW [$t(22)<1$, ns].

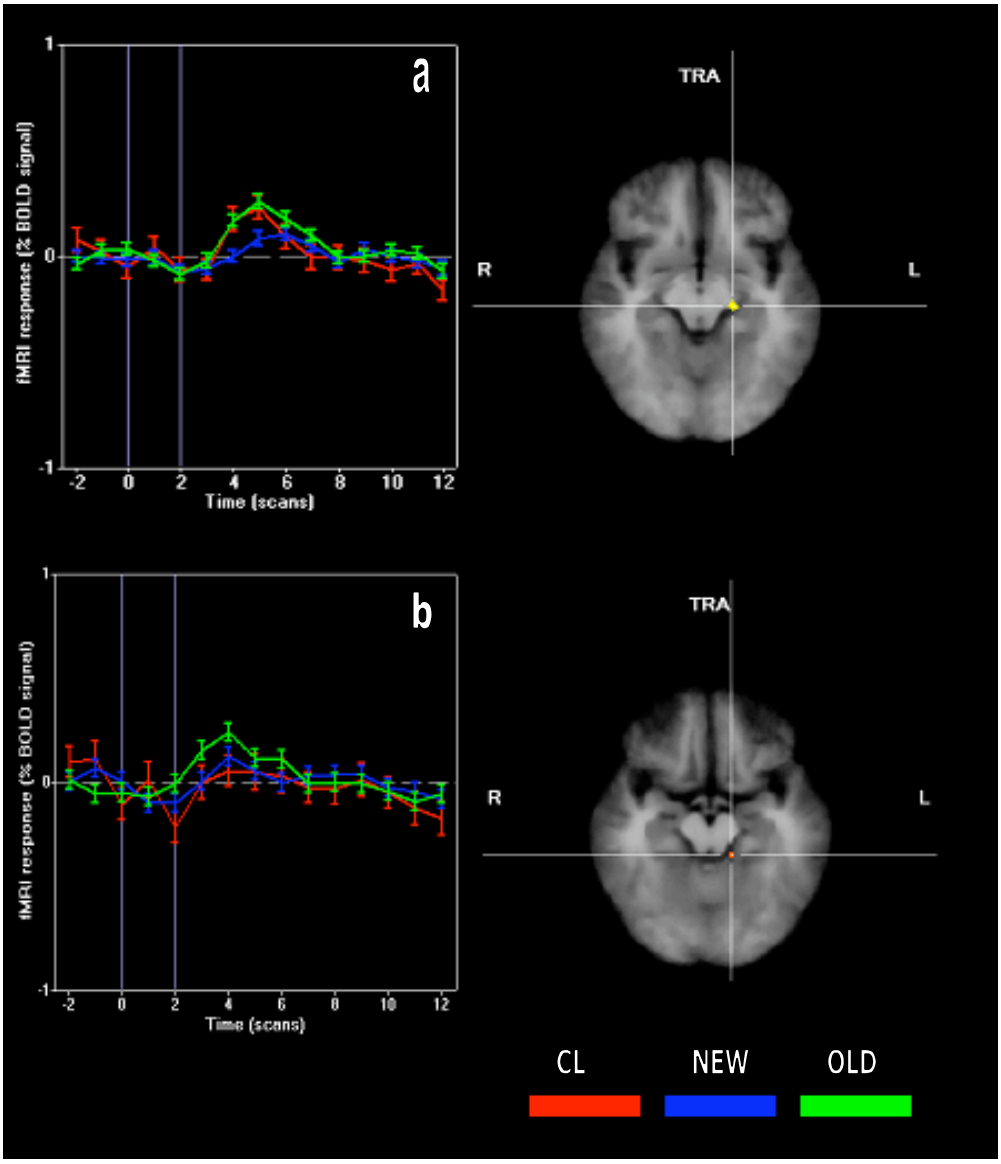

Figure 4 (a) *A left hippocampal region was more activated for OLD and CL words than for NEW words, with no difference between OLD and CL.* ***(b)*** *A left posterior parahippocampal region was more activated for OLD than for CL and NEW words, with no difference between CL and NEW.*

Source monitoring regions. Two left dorsolateral prefrontal cortex (DLPFC) regions and one right DLPFC region showed activity similar to the hippocampal region. All the three regions were more activated for OLD and CL words compared to NEW words. Pairwise contrasts of

the average BOLD signal yielded significant differences between CL and New [$t(22)= 3.84$, $p< 0.001$, $t(22)=4.17$, $p<0.001$ and $t(22)= 3.66$, $p< 0.001$, respectively] and between OLD and NEW [$t(22)=6.27$, $p< 0.001$, $t(22)=8.00$, $p<0.001$ and $t(22)= 4.70$, $p< 0.001$, respectively], but not between OLD and CL [$t(22) <1$, ns, $t(22)=1.49$, ns and $t(22)<1$, ns respectively].

Further, three regions within a middle frontal region indicated a third pattern of activation, interestingly showing a dissociation between right and left middle frontal gyrus (MFG). Two areas within right MFG differentiated between all the three wordtypes [CL vs. OLD, $t(22)= 4.78$, $p< 0.001$, $t(22)=4.85$, $p<0.001$, CL vs. NEW, $t\ (22)= 6.93$, $p< 0.001$, $t(22)=6.96$, $p<0.001$, and OLD vs. NEW, $t(22)= 3.04$, $p< 0.01$, $t(22)=2.98$, $p<0.01$, respectively] as shown in figure 5a and b, while one area within left MFG showed significant differences between CL and OLD words [$t(22)= 3.88$, $p< 0.001$], CL and NEW words [$t(22)= 5.16$, $p< 0.001$], but not between OLD and NEW words [$t(22)= 1.83$, $p<1$,ns].

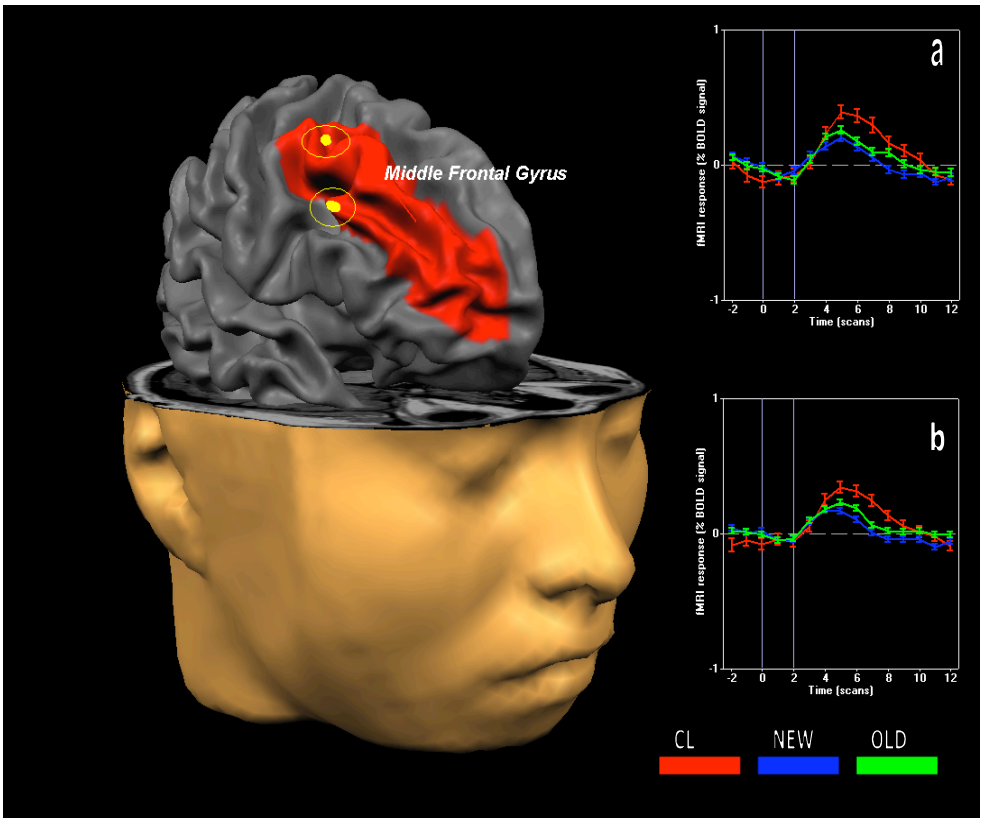

Figure 5 a & b. *Significant activations in the right MFG and their corresponding average BOLD signal. Both regions showed activation that differentiated all the three wordtypes, indicating enhanced activity for CL words.*

Wordtype by group effects

Nine separate pairwise contrasts were conducted within the previously defined ROIs (see table 3) in order to identify significant wordtype differences within and between the HFM and LFM group (HFM OLD vs. HFM NEW vs. HFM CL) vs. (LFM OLD vs. LFM NEW vs. LFM CL).

Medial temporal lobe. Pairwise contrasts of the average BOLD signal within left hippocampus yielded significant differences between LFM CL and HFM CL [$t(25)= 5.19$, $p< 0.001$] , between LFM OLD and HFM OLD [$t(25)= 5.47$, $p< 0.001$], and between LFM NEW and HFM NEW [$t(25)=8.89$, $p< 0.001$] as shown in figure 6, whereas pairwise contrast within each group for the same region revealed that the HFM group showed activation differences for OLD vs. NEW [$t(25)= 5.70$, $p<0.001$] and CL vs. NEW [$t(25)= 2.12$, $p<0.05$], but not for CL vs. OLD [$t(25)= 1.94$, ns], while the LFM group did not differ between any of the wordtypes [CL vs. OLD; $t(25)<1$, ns, CL vs. NEW; $t(25)<1$, ns, and OLD vs. NEW, $t(25)= 1.04$, ns].

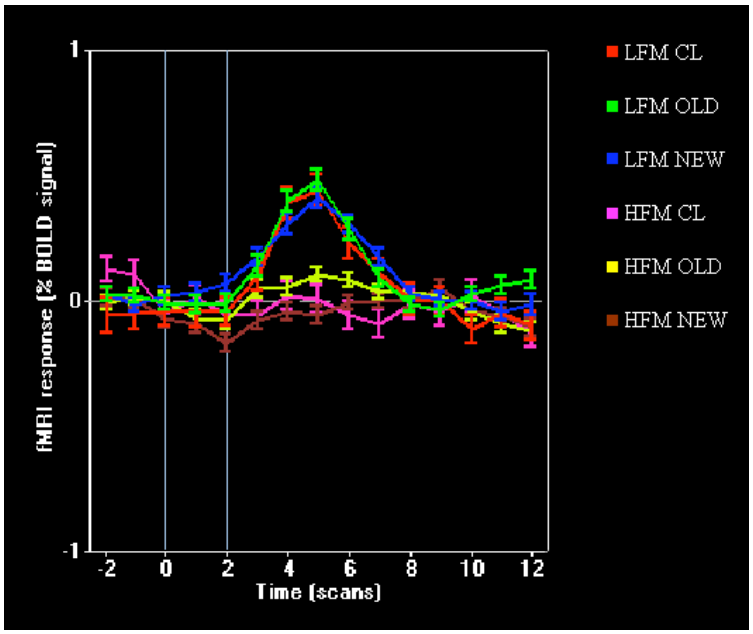

Figure 6. *The ROI within the left hippocampus showed significant group differences for all the wordtypes. The corresponding average BOLD signals for CL, OLD and NEW words for each group are shown.*

Pairwise contrasts of the average BOLD signal within left posterior parahippocampal gyrus yielded significant differences between LFM CL and HFM CL [$t(25)= 5.27$, $p< 0.001$], between LFM OLD and HFM OLD [$t\ (25)= 9.34$, $p< 0.001$], and between LFM NEW and HFM NEW [$t(25)=12.59$, $p< 0.001$]. Pairwise contrast within each group further revealed that

the LFM group showed activation differences for all the wordtypes [CL vs. OLD; $t(25)= 1.93$, $p<0.05$, CL vs. NEW; $t(25)= 3.90$, $p<0.001$, and OLD vs. NEW, $t(25)= 2.80$, $p<0.005$], while the HFM group did not [CL vs. OLD; $t(25)<1$, ns, CL vs. NEW; $t(25)<1$, ns, and OLD vs. NEW, $t(25)= 1.64$, ns].

Source monitoring regions. Activity within the right DLPFC ROI revealed that the LFM group differentiated between the three wordtypes, while the HFM group did not. Pairwise contrasts of the average BOLD signal yielded significant differences between LFM CL and HFM CL [$t(25)= 4.20$, $p< 0.001$], and between LFM NEW and HFM NEW[$t(25)= 2.88$, $p< 0.005$], but not between LFM OLD and HFM OLD [$t(25)=1.65$, ns]. Pairwise contrast for the three wordtypes within each group revealed that the LFM group differed between CL and NEW [$t(25)= 4.12$, $p< 0.001$], between OLD and CL [$t(25)= 2.37$, $p<0.05$] and between OLD and NEW [$t(25)= 2.49$, $p< 0.05$], while the HFM group differentiated between OLD and NEW [$t(25)= 4.160$, $p< 0.001$], but not between CL and NEW [$t(25)= 1.09$, ns], and not between CL and OLD [$t(25)=1.87$, ns]. A different pattern emerged within the two left DLPFC regions. None of the two groups differentiated CL words from OLD [LFM $t(25)<1$, ns; HFM $t(25)=1.76$, ns], whereas both groups differentiated OLD from NEW [LFM $t(25)=4.29$, $p< 0.001$; HFM $t(25)=4.58$, $p< 0.001$]. The LFM group also differentiated CL from NEW [$t(25)=3.98$, $p< 0.001$] while the HFM group did not [$t(25)=1.50$, ns].

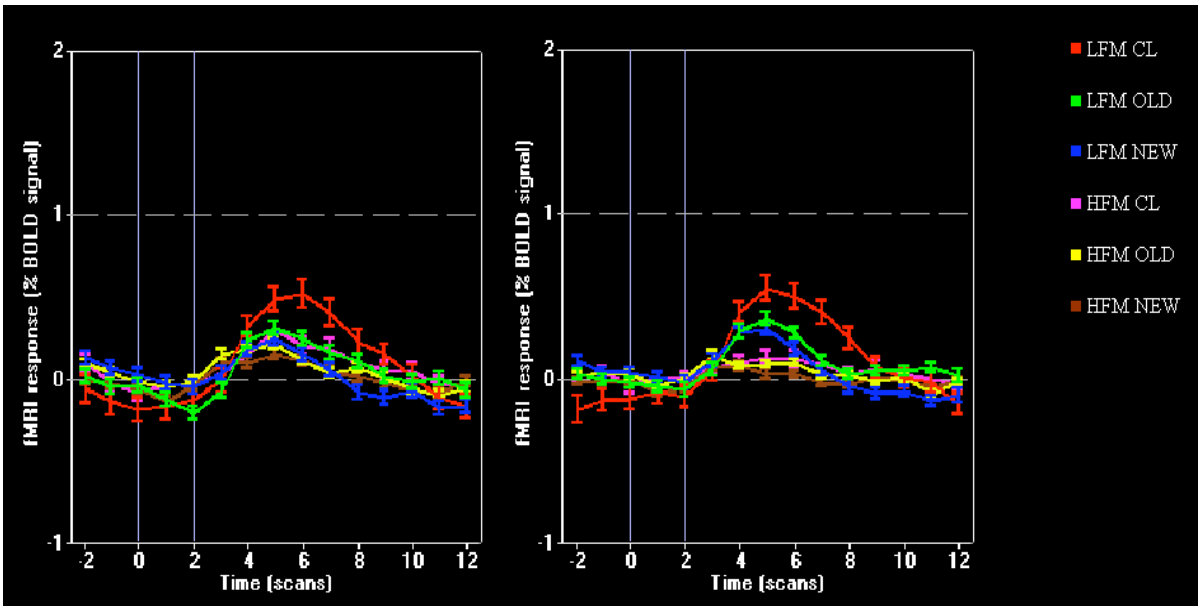

Figure 7. *The two ROIs within the right middle frontal gyrus showed significant differences for all the wordtypes when comparing the two groups. The corresponding average BOLD signals for CL, OLD and NEW words for each group are shown.*

The three regions within the middle frontal region again indicated a dissociation between right and left middle frontal gyrus (MFG). The two ROIs within right MFG showed significant group differences between LFM CL and HFM CL [$t(25)=8.73$, $p< 0.001$] and between LFM NEW and HFM NEW [$t(25)=7.714$, $p< 0.001$] and between LFM OLD and HFM OLD [$t(25)=7.27$, $p< 0.001$] as shown above in figure 7, while the left MFG only showed significant group differences for CL words [$t(25)=2.71$, $p< 0.01$] Interestingly, pairwise comparisons within groups showed that the LFM group differentiated between CL and OLD words within all the three regions [right, $t(25)=5.25$, $p< 0.001$, $t(25)=5.99$, $p< 0.001$ and left; $t(25)=3.90$, $p< 0.001$], while the HFM group did not [right, $t(25)=1.55$, ns, $t(25)<1$, ns and left; $t(25)=1.64$, ns].

A final pattern emerged within a medial frontal region, again indicating a dissociation between an anterior and a more posterior region within the superior frontal gyrus (SFG). The anterior region showed only significant group differences for NEW words [LFM NEW vs. HFM NEW, $t(25)=2.11$, $p<0.05$], while the more posterior region only showed significant differences for CL words [LFM CL vs. HFM CL, $t(25)=4.00$, $p< 0.001$]. Within group contrasts for both regions indicated that the LFM group showed significant activation differences for all the three wordtypes in both the anterior and posterior regions [CL vs. NEW, $t(25)=4.40$, $p< 0.001$, $t(25)=6.66$, $p< 0.001$; CL vs. OLD, $t(25)=2.30$, $p< 0.05$, $t(25)=3.80$, $p< 0.001$; and OLD vs. NEW, $t(25)=2.98$, $p< 0.005$, $t(25)=4.06$, $p< 0.001$, respectively], while the HFM group only showed activation differences between OLD and NEW words, but not between CL and OLD [OLD vs. NEW, $t(25)=2.29$, $p< 0.05$, $t(25)=3.37$, $p< 0.001$; CL vs. OLD, $t(25)<1$, ns, $t(25)<1$, ns respectively].

General recognition network differences

A final contrast between the high false memory (HFM) group and the low false memory (LFM) group was applied in order to identify regions that showed significant activation differences between the two groups during the recognition task. Significantly activated regions ($p<0.01$) showing enhanced activation for the LFM group were revealed in inferior parietal regions, frontal regions, occipital and temporal lobe regions (shown in figure 8a and b), this time revealing bilateral hippocampal activation (shown in figure 9).

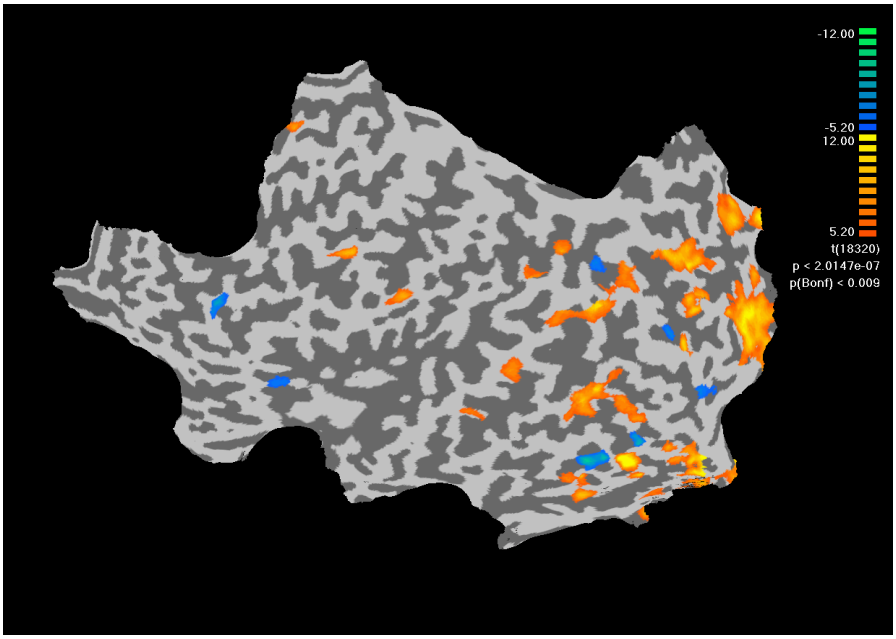

Figure 8a. *General recognition network differences between the LFM and HFM groups for the left hemisphere projected on a flatmap. Only voxels with a significant activity of p<0.01Bonferroni corrected for whole brain multiple comparisons are shown.*

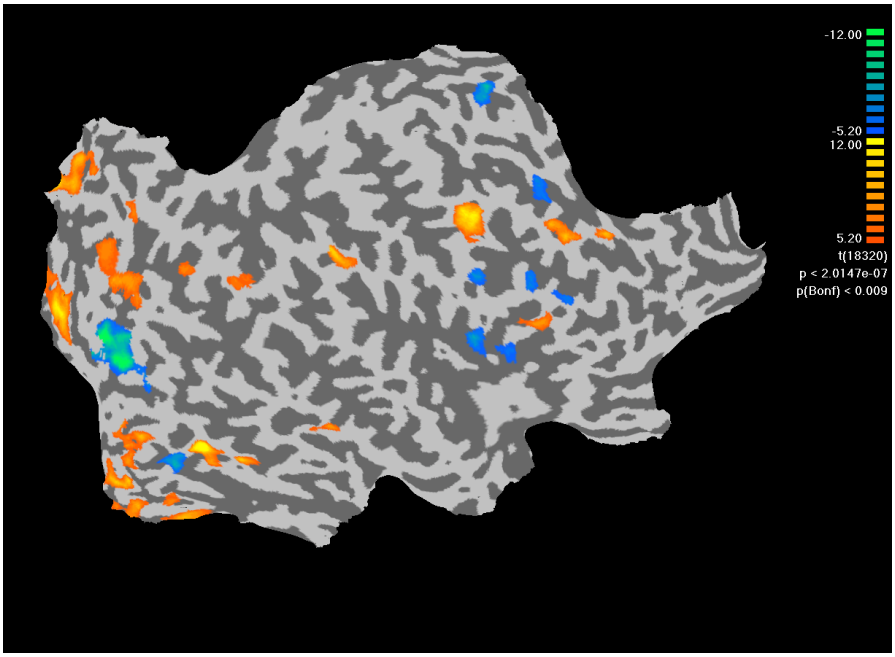

Figure 8b. *General recognition network differences between the LFM and HFM groups for the right hemisphere projected on a flatmap. Only voxels with a significant activity of p<0.01 Bonferroni corrected for whole brain multiple comparisons are shown.*

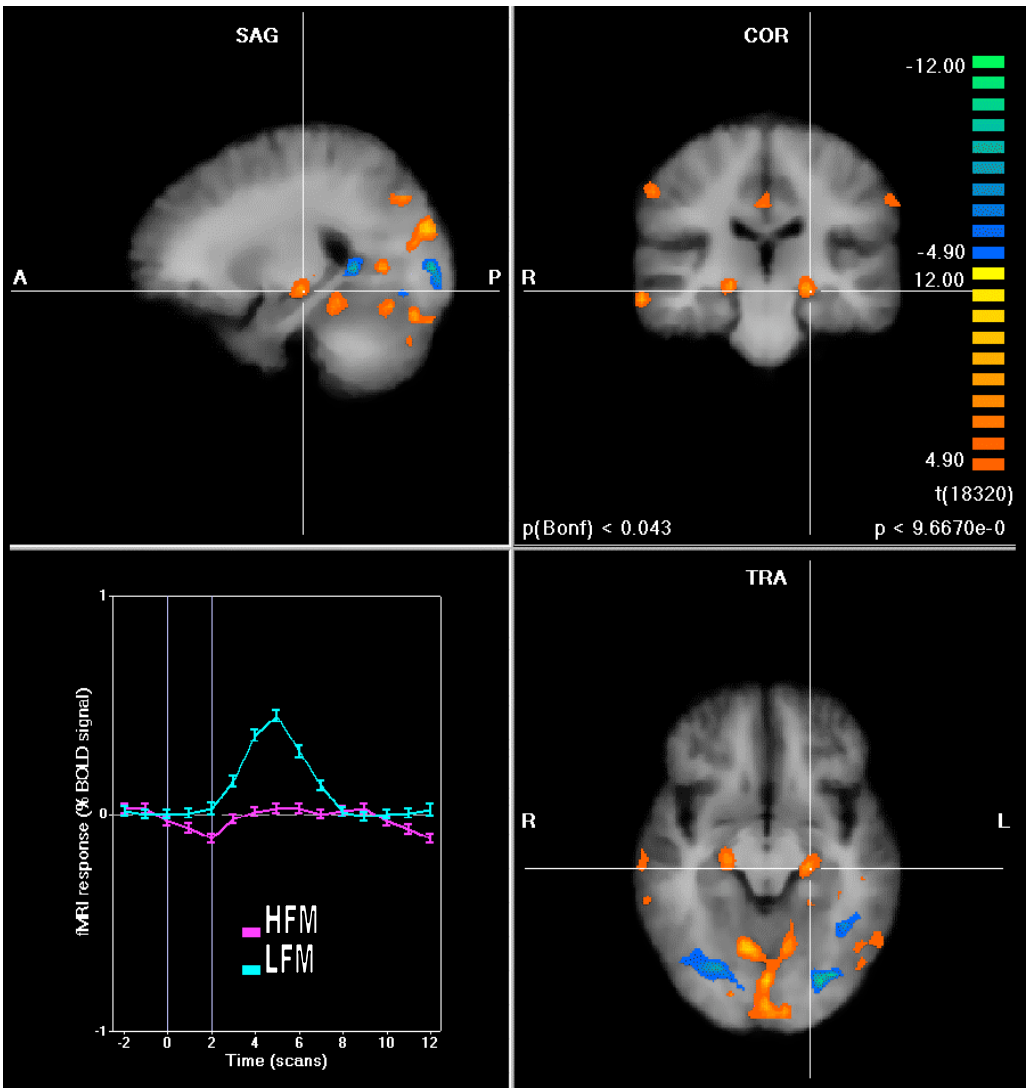

Figure 9. *Significant activations (p<0.05 Bonferroni corrected) in regions were the HFM and the LFM group differed. The cross indicates significant group differences within the left anterior hippocampus (HC) shown with the corresponding average BOLD signal for the two groups.*

Interestingly, the HFM vs. LFM contrast revealed strong group differences, specifically within left (-18,-21,-7) and right (20,-20,-5) anterior hippocampus and left posterior parahippocampal gyrus (-13,-36,-11)[2]. Both hippocampal regions and the left posterior parahippocampal region showed overall enhanced activation for the LFM group compared to the HFM group as shown in figure 10.

[2] The parahippocampal region described here included the posterior part of the parahippocampal gyrus, but extended inferior to our definition of the parahippocpal region.

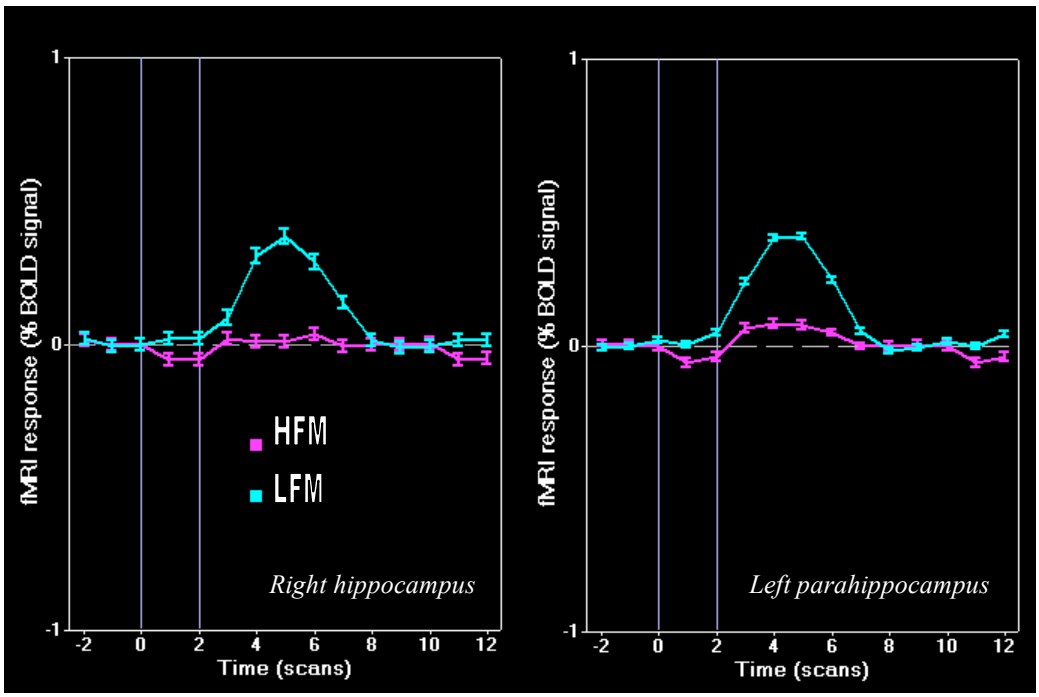

Figure 10. *Significant activation differences (p<0.05 Bonferroni corrected) within and left posterior parahippocampal gyrus were the HFM and the LFM group differed shown with the corresponding average BOLD signal for the two groups.*

Discussion

The overall level of accuracy observed in the present study reliably indicates that the participants understood and performed the encoding and recognition task in a satisfactory manner, while the observed strong false memory effect indicate that the presented stimuli induced false memories at rates comparable to previous studies (Stadler et al., 1999). Furthermore, distinct differences in accuracy, response time patterns and Remember/Know judgments were observed across the different trial types for each of the two groups. The HFM group's response pattern revealed high proportions of "remember" judgments and similar response times for critical lure words and previously studied words, while the LFM group's response pattern revealed low proportions of "remember" judgments and significantly different response times for critical lure words compared to previously studied words and new words. As explained in the introduction, false memories induced by the DRM paradigm are thought to occur due to a breakdown in source monitoring systems that differentiate the activation of internally generated concepts from the actual representations of previously studied words. This suggests that the LFM group showed an enhanced ability to successfully monitor the difference between the different wordtypes compared to the HFM group, more

often correctly rejecting the critical lure words. In the following paragraphs we will discuss the functional results related to these findings.

Medial temporal lobe

During retrieval, neural activity in the medial temporal lobe (MTL) has been found to increase as a function of the amount of information recovered, suggesting that hippocampus and the parahippocampal regions are involved in accurate remembering and recovery of stored memory traces (Wagner, Schacter, Rotte, Koutstaal, Maril, Dale, Rosen & Buckner, 1998). In line with this finding and previous false memory research (e.g. Cabeza et al., 2001; Schacter & Slotnick, 2004), the overall results of the present study revealed significant activations within the MTL. The left hippocampal region showed enhanced activation for previously studied words and critical lure words compared to new words, possibly reflecting recovery of semantic information, whereas the left posterior parahippocampal gyrus showed enhanced activation for previously studied words compared to critical lure words and new words, possibly reflecting recovery of perceptual or contextual information. The hippocampal pattern observed in the present study was also found in a study by Slotnick and Schacter (2004) in which abstract shapes were used. Since abstract shapes presumably have little semantic content, the authors suggested that hippocampus is activated by false memory for "gist" (Brainerd & Reyna, 2002), even if that gist is nonverbal.

When applying the group contrast, activations within bilateral anterior hippocampal regions and the left posterior parahippocampal gyrus were revealed. Interestingly, all three regions (left/right hippocampus and left parahippocampal gyrus) showed overall enhanced activation for the LFM group compared to the HFM group. Significant differences for each wordtype between the groups were also found in left hippocampus, whereas activations within left parahippocampal gyrus further revealed that the LFM group showed differential activation for all the wordtypes within this region, while the HFM group did not. These findings suggest that the LFM group recovered greater amounts of stored semantic information together with an enhanced retrieval of perceptual detail that enabled them to differentiate the wordtypes.

Source monitoring regions

Activity within prefrontal cortex has been found to be enhanced for previously studied words and critical lure words compared to new words, possibly reflecting monitoring of retrieved information (e.g. Cabeza, 2001). In line with this research, the three DLPFC regions observed in this study showed enhanced activation for previously studied words and

critical lure words compared to new words, indicating activity patterns similar to the hippocampal region. However, when applying the wordtype by group contrast, activity within the right DLPFC revealed that the LFM group differentiated between all the three wordtypes within this region, while the HFM group did not. Compared to the HFM group, the LFM group showed enhanced activity for the critical lure words, possibly reflecting the LFM group's ability to monitor the lack of recovered perceptual detail for the critical lure words. The HFM group only differentiated previously studied words from new words, possibly reflecting the HFM group's inability to monitor the small differences between previously studied words and critical lures, due to the low levels of perceptual detail recovered in parahippocampal gyrus. A different pattern emerged within the two left DLPFC regions. Neither of the two groups differentiated critical lure words from previously studied words, whereas both groups differentiated previously studied words from new words, possibly reflecting the processing of novel information.

Previous research has associated the middle frontal gyrus (MFG) with working memory and working memory load. A study by Fockert, Rees, Frith and Lavie (2001) revealed that several regions in the frontal cortex, including MFG, were active during conditions of high working memory load, possibly reflecting attentional efforts and the ability to avoid interfering information. In fact, a behavioural study by Watson, Bunting, Poole and Conway (2005) reported that individual differences in working memory capacity influenced individual susceptibility to false memories in the DRM paradigm. Their findings suggested that individual differences in working memory span influenced cognitive control and the ability to actively maintain task goals when presented with interfering information. In the present study, two regions within right middle frontal gyrus (MFG) differentiated between all the three wordtypes, while one area within left MFG differentiated critical lure words from previously studied words and critical lure words from new words. All three regions indicated increased activation for the critical lure words, possibly reflecting monitoring effort and working memory load. Further, the two right MFG regions revealed significant group differences for all three wordtypes showing enhanced activation for the LFM group, while the left region only showed significant group differences for critical lure words. Interestingly, pairwise comparisons within each group showed that the LFM group differentiated between critical lure words and previously studied words within all three regions, while the HFM group did not.

A similar pattern emerged within an anterior and a more posterior region of the left superior frontal gyrus (SFG) which is believed to be involved in higher levels of working

memory processing, such as monitoring and manipulation of information (e.g. Boisgueheneuc, Levy, Volle, Seassau, Duffau, Kinkingnehun, Samson, Zhang & Dubois, 2006). In the present study, the anterior region showed significant group differences for new words, while the more posterior region only showed significant group differences for critical lure words. Within group contrasts for both regions indicated that the LFM group differentiated between all the three wordtypes in both the anterior and posterior regions, while the HFM group only showed activation differences between previously studied words and new words. Taken together, our findings within MFG and SFG, indicated that the HFM group exhibits a reduced ability to monitor differences between the wordtypes compared to the LFM group, possibly reflecting reduced working memory capacity and reduced levels of cognitive control.

Encoding

According to previous research, a critical notion in the understanding of retrieval processes and their interaction with encoding processes is the acknowledgment of individual differences in encoding strategies (e.g. Hunt & Einstein, 1981; Hunt & McDaniel,1993) Since the findings in the present study indicated robust group differences during recognition within not only regions know to be involved in source monitoring, but also within regions involved in the storage and retrieval of episodic memory, we hypothesized that some of these differences had risen due to the use of different strategies in the encoding phase of the study. Interestingly, when analyzing the two groups' data separately, two distinct encoding networks were revealed as shown in figure 11.

Since we did not enquire about participants' encoding strategies, we can only speculate in terms of these results. Still, the strikingly different encoding networks displayed in figure 11, should indicate the use of different encoding strategies for the two groups. During encoding the HFM group showed an almost complete lack of hippocampal activity and decreased activity compared to baseline in several areas, especially in posterior parts of the MTL, including parahippocampal regions, while the LFM group showed strong activation in left hippocampal regions and posterior parts of MTL. Applying the group contrast revealed significantly enhanced activation for the LFM group in left hippocampus, bilateral posterior cingulate gyrus, bilateral inferior parietal regions, bilateral middle frontal regions, fusiform gyrus and bilateral posterior parahippocampal regions (see appendix for details).

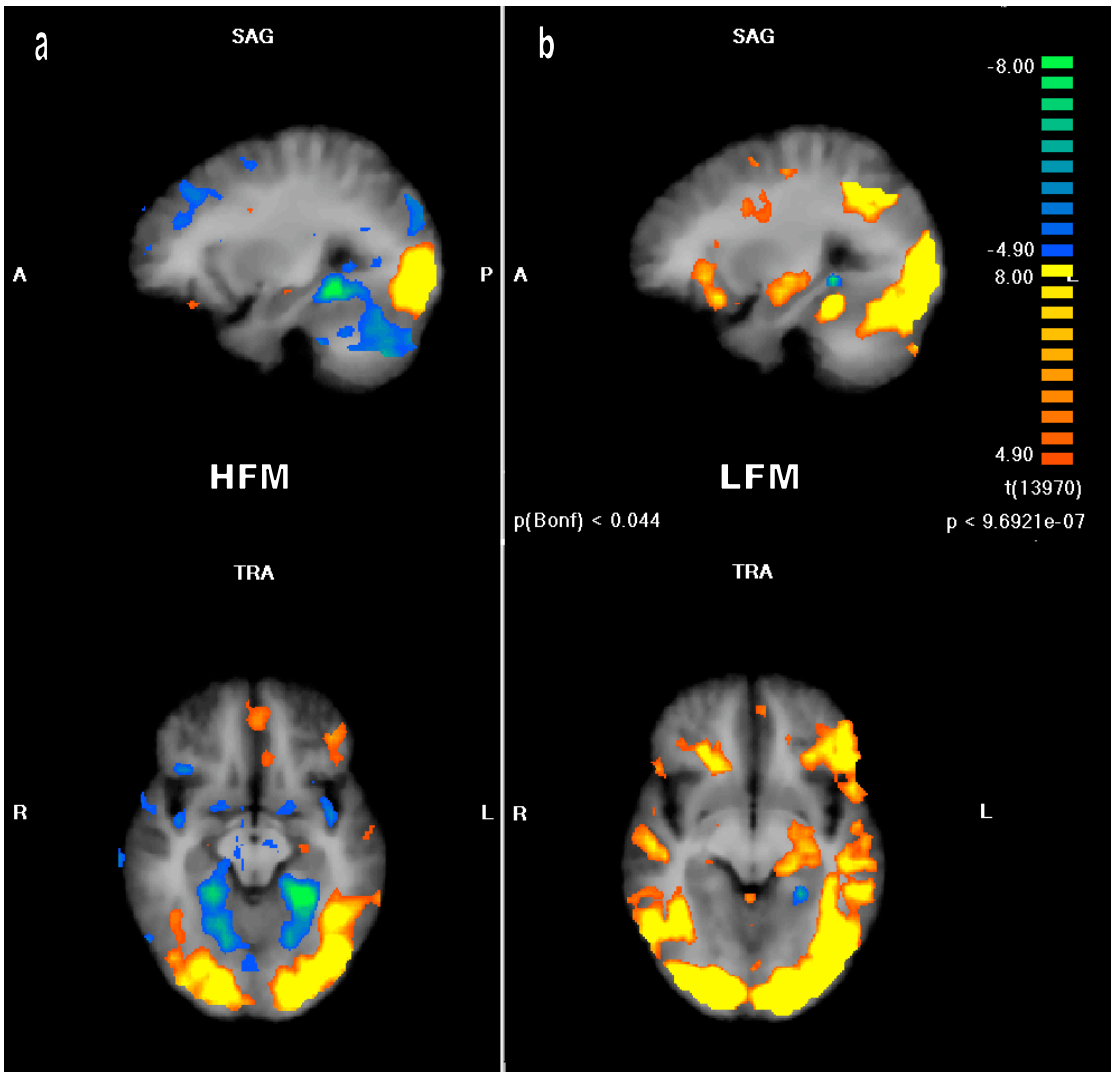

Figure 11. *Significant activations (p<0.05 Bonferroni corrected) when analyzing the two groups separately, showing (a) the encoding network for the HFM group and(b) the encoding network for the LFM group.*

Since we used a block design during the encoding runs, it is not possible to disentangle which encoding trial that induced a subsequent false memory. Still, Kim and Cabeza (2006) reported in their study on false and true memory formation that left prefrontal cortex was involved in both true and false memory formation, a finding consistent with the evidence that semantic elaboration, which has been associated with left PFC, tends to enhance both true and false remembering. According to Kim and Cabeza, encoding activity in the left MTL and early visual areas contribute mainly to true memory formation, whereas late visual areas are engaged in both true and false memory formation. These findings indicated that elaborative perceptual processing, but not basic sensory processing, contributes to false remembering. It is therefore possible to hypothesize that the participants in the HFM group used a form of

encoding strategy that incorporated low levels of perceptual information in parahippocampal regions, while elaborating semantic information storing only the gist of the memory. This, in turn, might have made the HFM group vulnerable to the strong sense of familiarity provoked by the critical lures during the recognition task

Limitations

Conventional fMRI studies usually implement the use of random effects in the analysis of the functional data. However, considering a sample size of ten participants and groups formed with five participants, the use of random effects was not recommended for the study. Instead a fixed effects analysis was used, possibly slightly increasing the chance of finding significant results within our specified groups. However, considering our strong hypotheses and the consistency of our findings, the results are not likely to be the byproduct of this statistical procedure.

In the introduction to this article we also mentioned that Daselaar and colleagues (2006) noted difficulties in identifying the different regions of interest within the MTL without the use of an unconventionally low statistical threshold. In fact, the authors reported that the anterior MTL regions they observed would not have survived the threshold conventionally used in event-related fMRI studies. Similar problems were encountered in the present study, and in order to identify reliable parahippocampal activity in the overall recognition network, we had to lower the threshold, showing non-significant voxels. Again, it is worth noting that different authors use different statistical criteria for "reliable activations" and that the Bonferroni corrected threshold of $p<0.05$ is quite conservative, especially when analyzing functional data within higher cognitive functions. Furthermore, the localization of our parahippocampal ROI was driven by a strong a priori hypothesis and theoretical frameworks which we argue justifies the use of this procedure.

Research within cognitive neuroscience has also reported that precise localization within MTL is rarely achieved, given the susceptibility-induced distortions associated with echo planar fMRI. In relation to this it is worth noting that fMRI methods are sensitive to magnetic field inhomogeneity, and in regions near air-filled sinuses such as the frontal lobes and the MTL, the magnetic field can vary greatly. Important voxels can therefore be shifted from their correct positions and cause geometric distortions which make it difficult to achieve accurate registration between activation maps and high resolution anatomical images (Hutton, Bork, Josephs, Deichmann, Ashburner & Turner, 2001). These problems cause uncertainty in

localization, especially for brain regions involved in higher cognitive function and could explain many of the inconsistent findings within the fMRI literature.

Regarding the lack of right hippocampal activity in the general recognition network presented in this article, an interesting point can be made here. In the Okado and Stark (2003) study the lack of hippocampal activity is explained by referring to confounding effects such as incidental encoding or the presence of episodic or source memory components in all the trials analyzed. According to the authors, such commonality may have resulted in similar levels of activity across the trial conditions in many of the MTL regions, ultimately flattening the activation pattern when applying the contrasts. In the present study only low levels of left hippocampal activity were found during recognition when examining the overall fMRI results. However, when applying the HFM vs. LFM group contrast a different picture was observed showing enhanced bilateral activation for the LFM group, indicating that high variability in MTL activity might better explain the lack of hippocampal activity reported in the Okado and Stark study.

Concluding remarks

In the present study we divided participants into high and low false memory groups based on their recognition accuracy for critical lure words in the DRM paradigm. Functional measurements revealed differences between the two groups demonstrating enhanced activation for the low false memory group within medial temporal lobe, specifically in hippocampal and parahippocampal regions, and in regions known to be involved in source monitoring processes, specifically in prefrontal regions. Within our functionally defined regions of interest, distinct group differences were observed in relation to the different wordtypes presented in the recognition task. The low false memory group reliably differentiated between the different wordtypes in the source monitoring regions, in some cases showing enhanced activation for the critical lure words, while the high false memory group did not show reliable activations that differentiated previously studied words from critical lure words in any of these regions, reflecting this group's inability to monitor the differences between previously studied words and critical lure words. The functional results further indicated that the low false memory group relied on monitoring of both semantic gist and stored perceptual information in order to perform the recognition task, while the high false memory group, in the absence of stored perceptual information, possibly only relied on semantic gist, making them highly vulnerable to the strong sense of familiarity provoked by the critical lure words.

Taken together, the findings of the present study suggest that neural signatures underlying individual differences in source monitoring abilities can be explored using fMRI measurements within frontal and medial temporal lobe regions, and that activation differences within these regions, to some extent, can explain individual differences in false memory susceptibility.

Future research

The false memory phenomenon and source monitoring differences observed in the present study and other previous studies hold important keys to future explorations of the mechanisms involved in normal memory function. Still, there are many obstacles to overcome. Even though block designed DRM paradigms reliably induce high rates of false memories which are convenient for high cost, time-limited fMRI studies, future research should develop and incorporate other paradigms than the DRM paradigm in order to broaden the investigation. With the above mentioned limitations in mind, paradigms should involve more "real" false memories than the highly associative critical lure words used here. Furthermore, future research should focus on the interaction between the encoding and retrieval stages, while incorporating measurements of variability in spatial localization and individual levels of memory performance in the analysis.

References

Brainerd, C.J., & Reyna, V.F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological science, 11,* 164-169.

Beck, A.T., & Steer, R.A. (1987). *Beck depression inventory: Manual.* San Antonio: Harcourt Brance Jovanovich.

Bernstein, D.P., Stein, J.A., Newcomb, M.D., Walker, E., Pogge, D., Ahluvalia, T., Stokes, J., Handelsman, L., Medrano M., Desmond, D., & Zule, W. (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse and Neglect, 27,* 169-190.

Bernstein, E.M., & Putnam, F.W. (1986). Development, reliability, and validity of a dissociation scale. *Journal of Nervous and Mental Disease, 174,* 727-735.

Boisgueheneuc, F., Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., Samson, Y., Zhang, S., & Dubois, B. (2006). Functions of the left superior frontal gyrus in humans: A lesion study. *Brain, 129,* 3315-3328.

Brédart, S. (2000). When false memories do not occur: Not thinking of the lure or remembering that it was not heard. *Memory, 8,* 125-128.

Brett, M., Johnsrude, I.S., & Owen, A.M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience 3,* 243-249.

Broadbent, D.E., Cooper, P.F., Fitzgerald, P., & Parkes, K.R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology, 21,* 1-16.

Cabeza, R., Rao, S.M., Wagner, A.D., Mayer, A.R., & Schacter, D.L. (2001). Can medial temporal lobe regions distinguish true from false? An event related functional MRI study of veridical and illusory recognition memory. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 4805–4810.

Daselaar, S.M, Fleck, M.S., & Cabeza, R. (2006). Triple dissociation in the medial temporal lobes: Recollection, familiarity and novelty. *Journal of neuropsychology, 96,* 1902-1911.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17-22.

Fockert, J.W., Rees, G., Frith, C.D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science, 291,* 1803-1806.

Gardiner, J.M., & Java, R.I. (1993). Recognition memory and awareness – an experimental approach. *European Journal of Cognitive Psychology, 5,* 337-346.

Geraerts, E., Smeets, E., Jelicic, M., van Heerden, J. & Merckelbach, H. (2005). Fantasy proneness, but not self-reported trauma is related to DRM performance of women reporting recovered memories of childhood sexual abuse, *Conciousness and cognition, 14,* 602-612.

Goebel, R., & Jansma, H. (2006). BrainVoyager QX Getting Started Guide. Brain Innovation, Maastricht, The Netherlands, Version 2.3 ,84-88.

Gonsalves, B., Kahn, I., Curran, T., Norman, K.A., & Wagner A.D. (2005). Memory strength and repetition suppression: Multimodal imaging of medial temporal cortical contributions to recognition. *Neuron, 47,* 751–761.

Henson. R. (2005). A mini-review of fMRI studies of human medial temporal lobe activity associated with recognition memory, *The Quarterly Journal of Experimental Psychology, 58B,* 340-360.

Hunt, R.R., & Einstein, G.O. (1981). Relational and item-specific information in memory. *Journal of verbal learning and verbal behaviour, 20*, 497-514.

Hunt, R.R., & McDaniel, M.A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32,* 421-445.

Hutton, C., Bork, A., Josephs, O., Deichman, R., Ashburner, J., & Turner, R. (2001). Image distortion correction in fMRI: A quantitative evaluation. *Neuroimage, 16,* 217-240.

Hyman , I.E.Jr., & Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory and Language, 35,* 101-117.

Kim, H., & Cabeza, R. (2006). Differential contributions of prefrontal, medial temporal, and sensory-perceptual regions to true and false memory formation. *Cerebral Cortex advance access* (2006, November 16). Retrieved December 14, 2006, from http://cercor.oxfordjournals.org/cgi/content/abstract/bhl122v1.

Lindsay, D.S., Hagen, L., Read, J.D., Wade, K.A., & Garry, M. (2004). True photographs and false memories. *Psychological Science, 15,* 149-154.

Loftus, E.F. (2003). Memories of things unseen. *Current Directions in Psychological Science, 13,* 145-147.

McDermott, K.B., & Watson, J.M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language, 4,* 160-176.

Merckelbach, H., Horselenberg, R., & Muris, P. (2001). The creative experiences questionnaire (CEQ): A brief self-report measure of fantasy proneness. *Personality and Individual differences, 31,* 987-995.

Okado, Y., and Stark, C. (2003). Neural processing associated with true and false retrieval. *Cognitive, Affective, and Behavioral Neuroscience, 3,* 323–334.

Roediger, H.L., III, and McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 803–814.

Schacter, D.L., Norman, K.A., and Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49,* 289-318.

Schacter, D.L., Slotnick, S. (2004). The cognitive neuroscience of memory distortion. *Neuron, 44,* 149–160.

Schacter, D.L, Verfaellie, M., & Pradere, D. (1996). Neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory and Language, 35,* 319-334.

Slotnick, S.D., and Schacter, D.L. (2004). A sensory signature that distinguishes true from false memories. *Nature Neuroscience, 7,* 664–672.

Stadler, M.A., Roediger, H.L., & McDermott, K.B. (1999). Norms of word lists that create false memories. *Memory and cognition, 27,* 494-500.

Talaraich, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: An approach to cerebral imaging*. Stuttgart: Thieme.

Tulving, E. (1958). Memory and consciousness. *Canadian Psychology - Psychologie Canadienne, 26,* 1-12.

Wagner, A.D., Schacter, D.L., Rotte, M., Koutstaal, W., Maril, A., Dale, A.M., Rosen, B.R., Buckner, R.L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science, 281,* 1188-1191.

Watson, J.M., Bunting, M.F., Poole, B.J., & Conway, A.R.A. (2005). Individual differences in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology, 31,* 76-85.

Appendix A

The 26 criticallures used in the study (Dutch and English);

1. RIVIER	RIVER
2. ZOET	SWEET
3. STOEL	CHAIR
4. RUW	ROUGH
5. STAD	STATE
6. FRUIT	FRUIT
7. VUILNIS	GARBAGE
8. ROOK	SMOKE
9. BOOS	MAD
10. DROEFHEID	SADNESS
11. SPIJT	REGRETT
12. MOORD	MURDER
13. VOET	FOOT
14. BROOD	BREAD
15. MUZIEK	MUSIC
16. MISHANDELING	ABUSE
17. NAALD	NEEDLE
18. MAN	MAN
19. DIEF	THIEF
20. KOUD	COLD
21. AUTO	CAR
22. ANGST	ANXIETY
23. WOEDE	ANGER
24. JALOEZIE	JEALOUSY
25. TWIJFEL	DOUBT
26. WANHOOP	HOPELESSNESS

Appendix B

Encoding differences when applying the HFM vs. LFM contrast

(p<0.001, Bonferroni corrected)